Scene integration without awareness: No conclusive evidence for processing scene congruency during continuous flash suppression

Pieter Moors, David Boelens, Jaana van Overwalle, Johan Wagemans

Laboratory of Experimental Psychology, Department of Brain & Cognition, University of Leuven (KU Leuven), Leuven, Belgium

Correspondence to:

Pieter Moors

Laboratory of Experimental Psychology

Department of Brain & Cognition

University of Leuven (KU Leuven)

Tiensestraat 102, box 3711

B-3000 Leuven

Belgium

Pieter.Moors@ppw.kuleuven.be

+32(0)16325983

ABSTRACT

A recent study showed that scenes with a semantically incongruent object-background relationship break interocular suppression faster than scenes with a semantically congruent relationship. These results implied that semantic relations between objects and background of a scene could be extracted in the absence of visual awareness of the stimulus. In this study, we assessed the replicability of this finding and tried to rule out an alternative explanation based on low-level differences between the stimuli. Furthermore, we used a Bayesian analysis to quantify the evidence in favor of the presence or absence of a scene congruency effect. Across three experiments, we found no convincing evidence for a scene congruency effect nor a modulation of scene congruency by scene inversion. These findings question the generalizability of previous observations and cast doubt on whether genuine semantic processing of object-background relationships in scenes can manifest during interocular suppression.

Understanding the scope and limits of unconscious visual processing has become a central research topic in cognitive neuroscience (Dehaene & Changeux, 2011). A recent study claimed to have obtained evidence that complex, high-level visual scene processing can happen unconsciously (Mudrik, Breska, Lamy, & Deouell, 2011). These authors presented participants with scenes that were rendered invisible through continuous flash suppression (CFS, Tsuchiya & Koch, 2005). Scene congruency was manipulated and participants had to indicate when an initially suppressed scene broke suppression. Mudrik et al. (2011) observed that incongruent scenes broke suppression faster than congruent scenes. This led the authors to argue that consciousness of a scene is not required for high-level scene processing mechanisms to unfold and hence extract the congruency relation between object and background.

Because the results of Mudrik et al. (2011) have profound implications for theories on the extent of unconscious visual processing during CFS, it is of utmost importance that the congruency effect can be attributed to genuine scene processing mechanisms rather than differences between scenes on low-level visual aspects. Although image analyses (Itti & Koch, 2000; Neumann & Gegenfurtner, 2006) on the stimulus set used in these experiments did not seem to reveal any consistent bias on low-level visual aspects for the (in)congruent category, a stronger control for image-related characteristics is to include an experimental condition in which the scenes are inverted, which dramatically reduces their identifiability while fully preserving the low-level image properties.

The goal of this study was threefold. First, we wanted to assess the replicability of the original findings. Second, by including a scene inversion condition, we wanted to rule out any potential low-level confounds related to the particular stimulus set. Third, we complement the traditional repeated measures ANOVA with a Bayesian analysis based on linear mixed-effects modeling with crossed random effects for participants and stimuli (Clark, 1973). Given that the experiment consists of presenting various exemplars of congruent and incongruent scenes to participants, a random effect of stimulus should also be included if one aims to generalize to the population of congruent and incongruent scenes. Furthermore, Bayesian statistics allows to quantify the evidence for the *absence* of an effect (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

EXPERIMENT 1

Experiment 1 consisted of a replication experiment of Mudrik et al. (2011), using the same stimuli and methods, yet also including a scene inversion condition. If the results observed in Mudrik et al. (2011) are genuinely attributable to unconscious scene processing, we predicted to observe a scene congruency effect in the upright, but not in the inverted condition. Conversely, if the effect does not pertain to processing the semantic aspects of the interocularly suppressed scenes, we predict a similar congruency effect in the inverted condition and, critically, no (statistical) interaction between scene inversion and scene congruency.

METHODS

Participants

45 people participated in the study in return for money or course credit. All participants were naïve with respect to the purposes of the study and had normal or corrected-to-normal vision. Because the original sample size of 18 participants used in Mudrik et al. (2011) yielded a post-hoc power of 75%, we decided to substantially increase the sample such that the power based on the effect size reported in the original study was 99% for this experiment. This increase in sample size was further motivated by the fact that one needs sufficient measurements for each item to fit a linear mixed-effects model with crossed random effects.

Apparatus

Stimuli were shown on two 19.8-in. Sony Trinitron GDM F500-R (2048 x 1536 pixels at 60 Hz, for each) monitors driven by a DELL Precision T3400 computer with an Intel Core Quad CPU Q9300 2.5 GHz processor running on Windows XP. Binocular presentation was achieved by a custom made stereo set-up. Two CRT monitors, which stood opposite to each other (distance of 220 cm), projected to the left and right eye respectively via two mirrors placed at a distance of 110 cm from the screen. A head- and chin rest (15 cm from the mirrors) was used to stabilize fixation. The effective viewing distance was 125 cm. Stimulus presentation, timing and keyboard responses were controlled with custom software programmed in Python using the PsychoPy library (Peirce, 2007, 2009).

Stimuli

The background of the display consisted of a random checkerboard pattern to achieve stable binocular fusion. The size of the individual elements of the checkerboard was equal to 0.34°. In both eyes, a gray frame (10° by 10°) was superimposed on the checkerboard pattern to present the stimuli. A black (eye dominance measurement) or white (main experiment) fixation cross was continuously present during the experiment (size 0.5 by 0.5°). In the eye dominance measurement phase, the target consisted of an arrow (maximal width 4°, maximal height 2°) and the CFS mask consisted of 150 gray squares with randomly picked sizes between 1 and 2° and a random luminance value for each element.

The scene stimuli (2.86° x 2.03°) were the same as in Mudrik et al. (2011). For a detailed description of the stimulus set, we refer to the original study. In short, the scenes depicted various human actions involving a certain object. Both congruent and incongruent versions of the scene were created by pasting an object into the scene that was related or unrelated to the action (Mudrik et al., 2011). In the main experiment, the CFS mask consisted of 200 square elements with a randomly chosen color and size between 0.75° and 1.5° for each element. The positions of the elements were generated in a 5.26° x 5.26° square window centered at fixation. Because the maximal size of each element was 1.5°, the effective size of the CFS mask was thus minimally 5.26° x 5.26° and maximally

6.76° x 6.76° (compared to always 5.26° x 5.26° in the original study). In both the eye dominance phase as well as the main experiment, the refresh rate of the CFS mask was set at 10 Hz.

Procedure

In the first part of the experiment, participants' eye dominance was measured using the method of Yang, Blake, and McDonald (2010). On every trial, a fixation cross was presented for 1 second. Next, an arrow that gradually increased to 100% contrast in 2 seconds was presented to one eye and the CFS mask in the other. Upon breakthrough of the arrow, participants had to indicate the direction of an arrow by pressing "1" or "3" on a numerical keyboard for left and right pointing directions, respectively. The CFS mask was randomly presented to the left or right eye (40 trials per eye) for a total of 80 trials. The dominant eye was determined as the eye in which mean suppression time was shortest when the arrow was presented to that eye. In all subsequent phases of the experiment, the CFS mask should have been presented to the participants' dominant eye. Due to a programming error, however, the mask was always presented to the participant's right eye (in Experiments 1 and 2). In this experiment, 44% of the observers was left-, rather than right-eye dominant according to the criterion defined above , and thus received the mask in their non-dominant (right) eye.

The main experiment consisted of 160 CFS trials, divided into four blocks of 40 trials. In each block, all four conditions (all combinations of scene congruency and scene inversion) were balanced. Only one version of a scene was presented in each block. Within each block, the ordering of conditions was completely randomized, whereas in the original study a constraint was used that items of the same type could not be presented on four or more consecutive trials. Before the start of the main experiment, participants completed 16 practice trials based on four scenes that were not included in the main experiment.

On each trial, a CFS mask was presented to the participants' right eye while the scene was presented to the left eye and gradually increased from 0 to 100% contrast in steps of 10% every 100 ms. After the scene had reached full contrast, the CFS mask began to decrease in contrast to 0% over the course of 5.1 seconds. Upon breakthrough, participants had to indicate as fast as possible whether the scene was presented to the left or right of fixation (see Figure 1 for an overview of the trial sequence). Contrary to the original study, we did not perform a post-experimental rating session in which participants were asked to categorize all scene stimuli as being unusual or not. In the original study, incorrectly categorized stimuli were then removed prior to the start of the analysis. Because categorization performance could influence the results, we invited participants afterwards to perform the rating session in an online experiment. Here, we presented the scene stimuli in a random order and participants had to indicate for each stimulus whether it was unusual or not.

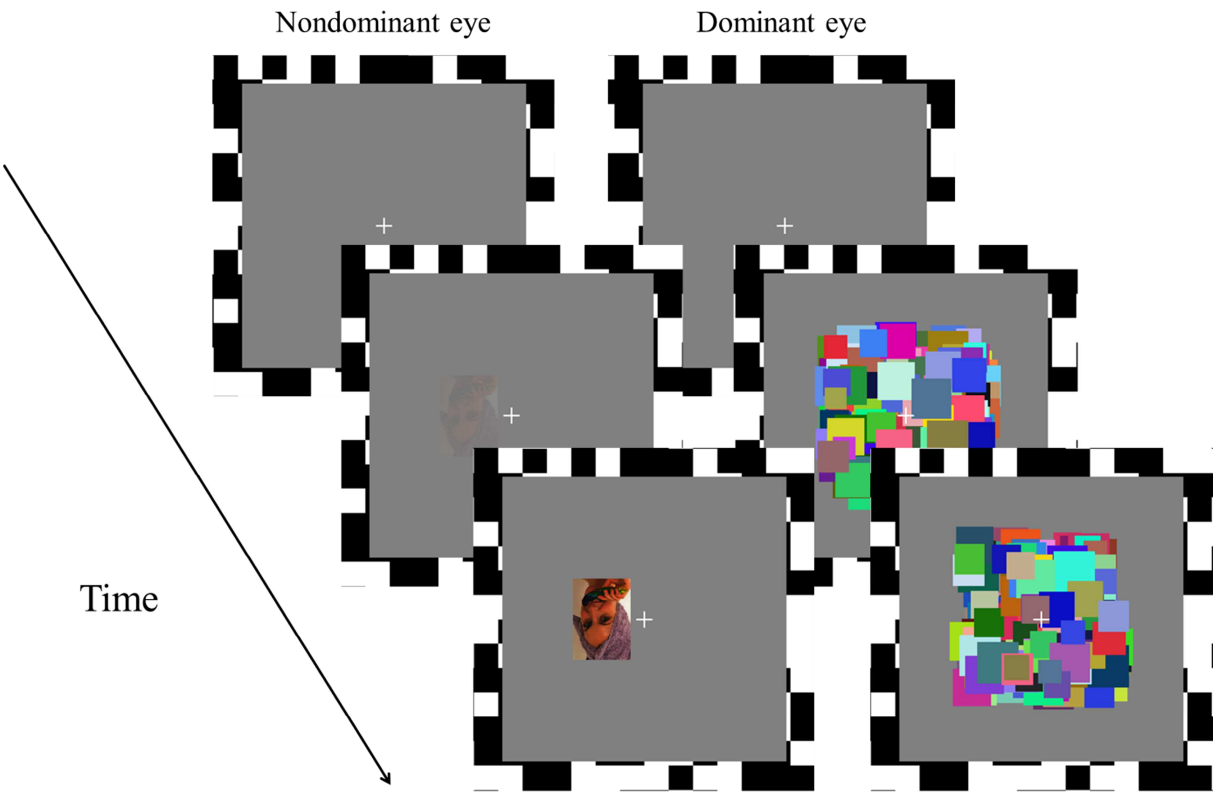

Figure 1. Trial sequence for all experiments. A fixation cross was presented to both eyes for 1 second, after which the scene stimulus was presented to the nondominant eye and the CFS mask to the dominant eye. The scene stimulus gradually increased in contrast for 1 second after which the CFS mask started decreasing in contrast for 5.1 seconds. Upon breakthrough, participants had to indicate as quickly as possible whether the scene stimulus was presented to the left or the right of the fixation cross.

RESULTS

All analyses were performed on the correct trials (*M* = 0.99, *SD* = 0.01) of which the suppression time did not exceed the time at which the mask reached 0% contrast (*M* = 0.92, SD = 0.13), after removing outliers defined as suppression times higher than the mean suppression time plus three times the standard deviation (for each observer separately, *M* = 0.005, *SD* = 0.007) . For the Bayesian analysis, we logarithmically transformed the suppression times to account for the positive skew in the suppression time distributions. Bayes Factors (BF) were computed to quantify the evidence for the presence/absence of a main effect or interaction. The R package BayesFactor (version 0.9.11-1, default settings, "medium" prior scale for fixed effects and "nuisance" prior scale for random effects) was used to compute the BFs (Morey & Rouder, 2015). All models were linear mixed-effects models with crossed random effects, including a random intercept for both participants and stimuli (Rouder, Morey, Speckman, & Province, 2012). To compute the BFs for the main effects and interaction, we compared a full model (including the two main effects and interaction) with a reduced model in which the effect of interest was not included (i.e., similar to the classical repeated measures ANOVA). According to the classification provided by Jeffreys (1961) a BF of 3 constitutes substantial evidence for one model over the other whereas a BF of 10 is considered to be strong evidence for one model over the other. It is important to stress that Bayes Factors constitute a relative measure of evidence for one statistical model compared to another (i.e., it is a ratio of marginal likelihoods computed for two statistical models). In this paper, we always report Bayes Factors with the reduced model in the numerator and the full model in the denominator. Thus, BFs > 3 indicate evidence for the *absence* of a main effect or interaction under consideration. BFs < 0.3 indicate evidence for the *presence* of the effect under consideration. Because the Bayes Factor is asymmetric around 1, we visualize the Bayes Factors after logarithmically transforming them (with base 10 logarithm) such that a BF of 1 or -1 indicates strong evidence for the absence or presence of an effect, respectively.

Before analyzing the data, we first checked the consistency across observers with respect to which stimuli broke suppression fast and slow. That is, if the images were being processed during suppression, at least the low-level image characteristics should have an influence on suppression times and this should be apparent from calculating Cronbach's α for the suppression times. Indeed, Cronbach's α was high for the suppression durations (α= .83) indicating that there was consistency across observers in which stimuli broke suppression fast and slow.

A two-way repeated measures ANOVA on the mean correct suppression times revealed no main effect of scene congruency (congruent scene: M = 2.53, SD = 0.88; incongruent scene: M = 2.56, SD = 0.86), $F(1, 44) = 1.08$, $p = .30$, $d = -0.11$, a main effect of scene inversion (upright scene: M = 2.49, SD = 0.87; inverted scene: M = 2.61, SD = 0.86), $F(1, 44) = 16.1$, $p = .0002$, $d = -0.44$, and no interaction between scene congruency and scene inversion, $F(1, 44) = 0.022$, $p = .88$, $d = -0.02$. As is apparent from Figure 2, inverted scenes on average broke suppression slower than upright scenes. Furthermore, congruent scenes on average broke suppression numerically faster than incongruent scenes, yet the p-value associated with this difference did not reach the significance threshold.

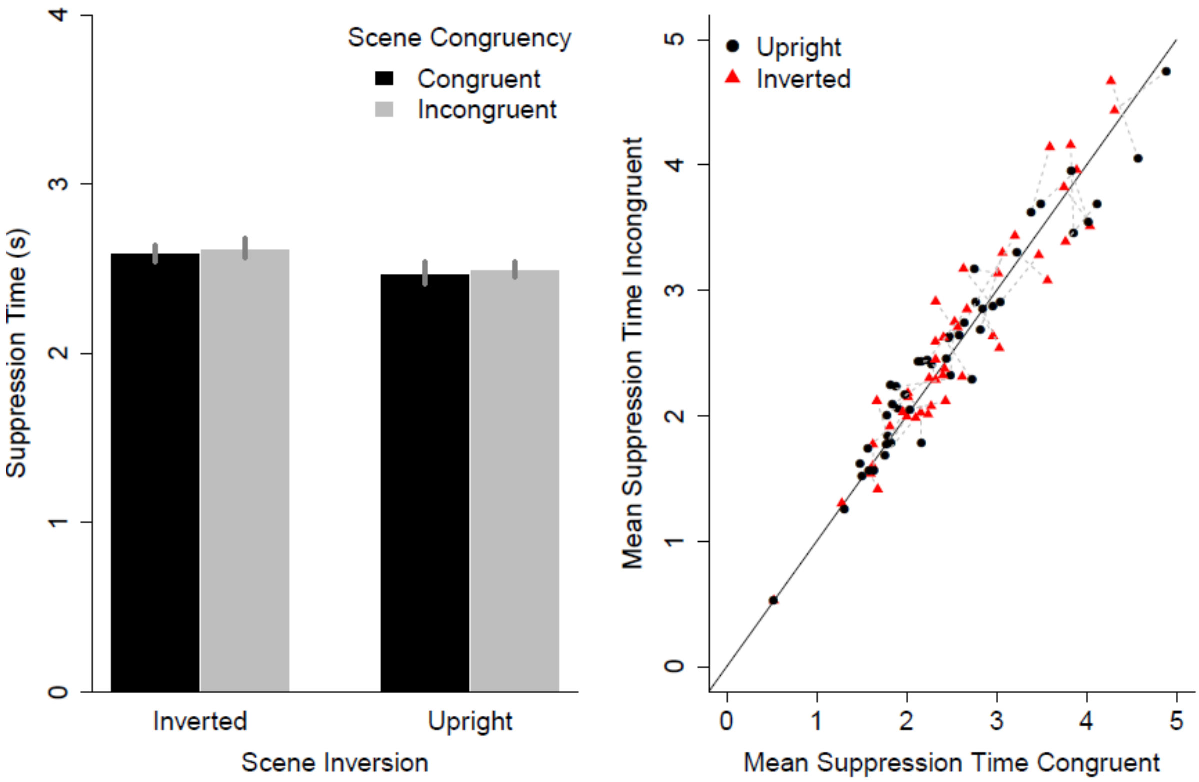

Figure 2. Results of Experiment 1. (left) Bar plot depicting mean suppression times for each condition. Error bars refer to 95% within-subject confidence intervals with the adjustment suggested by Morey (2008). (right) Scatter plot of mean suppression times for congruent and incongruent scenes for each participant. Black dots refer to the upright condition, red triangles to the inverted. The gray dashed lines connect the same participants.

The results of the repeated measures ANOVA were complemented by those of the Bayesian analysis. Specifically, the Bayes factor indicated strong evidence in favor of the absence of a scene congruency effect (BF = 17), strong evidence in favor of the presence of a scene inversion effect (BF < .01) and strong evidence for the absence of an interaction effect (BF = 24).

Figure 3 depicts a sequential analysis of the data for both statistical techniques. The upper panels depict how the p-values of the repeated measures ANOVA evolved as more participants were tested. As is apparent, several times throughout data collection, the main effect of congruency passed the significance threshold of .05. In contrast, the interaction never reached significance. In this respect, it is interesting to compare these patterns to the evolution of the Bayes Factors, which depict a different story. Here, the main effect of congruency nearly always hovers around a value of log10BF = 1 (i.e., a BF of 10), indicating strong evidence for the absence of a congruency effect. The BF of the interaction effect also shows a gradual increase towards more evidence in favor of the absence of an interaction effect. Interestingly, for the main effect of inversion, both patterns show a gradual increase in evidence for an inversion effect.

Taking into account participants' categorization on the rating experiment did not change the results of this experiment (see Supplementary Materials). We also reanalyzed the data by excluding all participants for whom the CFS mask was presented to the non-dominant eye. Again, this did not influence the results (see Supplementary Materials).

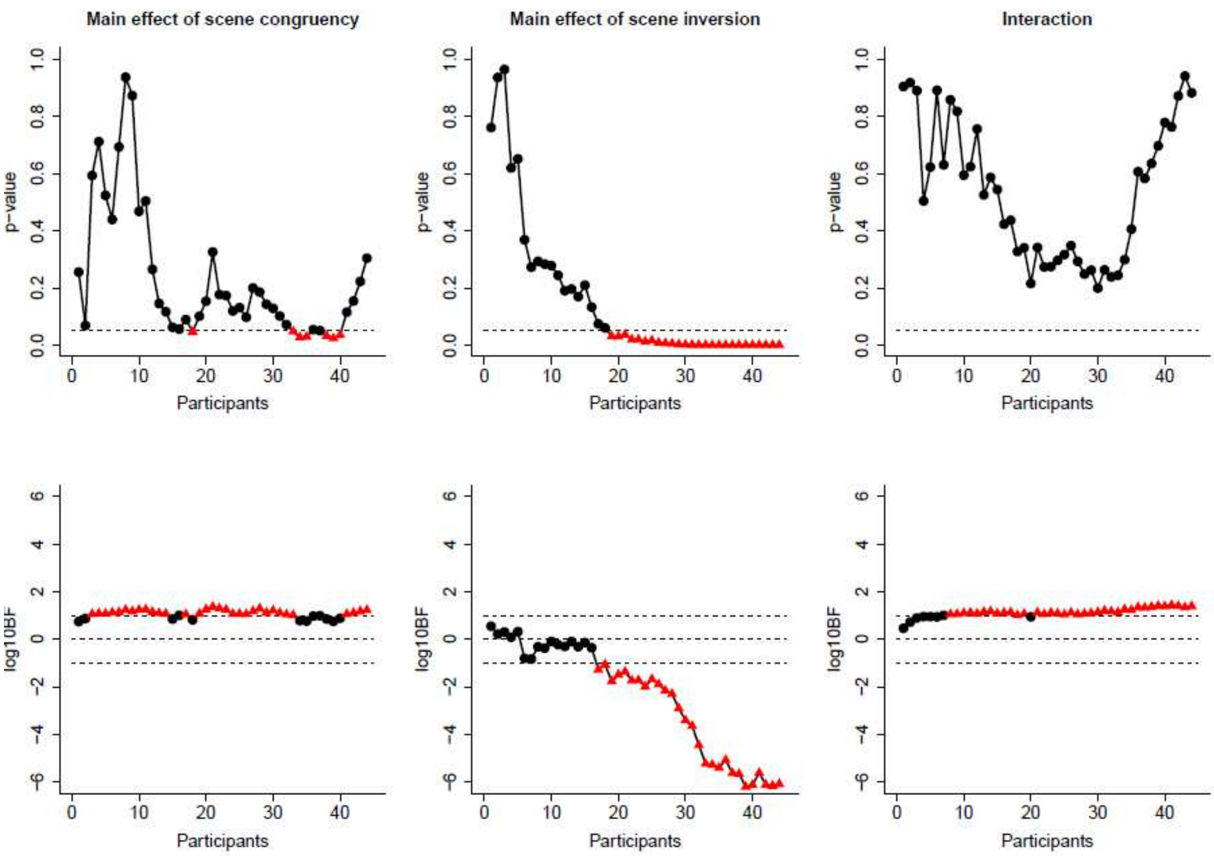

Figure 3. Sequential analysis of the data. (top row) Evolution of p-values as more data were collected for the main effects of scene congruency and scene inversion and their interaction, respectively. Black dots indicate p-values higher than .05 whereas red dots indicate p-values smaller than .05. (lower row) Evolution of Bayes Factors. Bayes Factors higher than 0 indicate evidence for the absence of an effect whereas Bayes Factors smaller than 0 indicate evidence for the presence of an effect.

DISCUSSION

The goal of Experiment 1 was to replicate the scene congruency effect observed in Mudrik, Breska, et al. (2011) and to assess whether it would be influenced by scene inversion. If genuine integration between the objects of a scene and its background can still proceed during CFS, one would predict that inverting the scenes would reduce or diminish the scene congruency effect. The results of Experiment 1 showed no convincing evidence for a main effect of scene congruency nor an interaction between scene congruency and scene inversion. In contrast, a reliable scene inversion effect was found in that upright scenes broke suppression faster than inverted scenes.

To increase our confidence in the absence of a scene congruency effect, we ran a second experiment in which we increased the number of trials for each observer. Furthermore, we dropped the inversion condition. This allowed us to run a quasi-exact replication of the original study and to assess whether increasing the number of trials for each observer would give us more power to detect a scene congruency effect. Furthermore, we decided to drop the mask fade-out procedure because this forced us to exclude a high number of trials for observers for which suppression was very effective.

EXPERIMENT 2

METHODS

Participants

24 new people participated in the study in return for money or course credit. All participants were naïve with respect to the purposes of the study and had normal or corrected-to-normal vision.

Apparatus and Stimuli

The set-up and stimuli were exactly the same as in Experiment 1. Inverted scenes were no longer included.

Procedure

The procedure was the same as in Experiment 1 except as noted here. In this experiment, 50% of participants was identified as left-eye dominant. The CFS mask was thus presented to the non-dominant eye for half of the participants. We did no longer include the scene inversion condition. The experiment was set up such that each block constituted a fairly exact replication of the original study. That is, the experimental procedure was the same as in the original study and this procedure was repeated four times. We only did no longer include the mask fade-out procedure to ensure that the absence of a scene congruency effect could not be attributed by limiting the observations for observers for which suppression was very effective and thus excluding the upper tail of the suppression time distributions for these participants. Experiment 2 also did not include the post-experimental rating session, but we contacted the participants afterwards to complete it online, as we did for Experiment 1.

Design

Scene congruency was the only factor that was manipulated. Participants completed 8 practice trials, and the main experiment consisted of four repetitions of the full stimulus set amounting to 320 trials in total.

RESULTS

The results were analyzed in the same way as in Experiment 1. That is, all analyses were performed on the correct trials (first block: M = 0.98, SD = 0.02; all data: M = 0.99, SD = 0.02), after removing outliers defined as suppression times higher than the mean suppression time plus three times the standard deviation (for observers separately, first block: M = 0.007, SD = 0.01; all data: M = 0.02, SD = 0.008). We separately report analyses for the first block only (quasi-exact replication of the original study) and all four blocks combined.

First, we again analyzed the consistency of suppression times for images across observers. This again indicated high reliability for the suppression times of the images across observers (α = 0.67 and 0.81, for the first block and all data respectively). Interestingly, the ordering of which items broke suppression fast and slow correlated highly across both experiments (r = 0.89, BF < 0.01).

As in Experiment 1, no effect of scene congruency was observed in either the first block (congruent scenes: M = 3.06, SD = 1.16; incongruent scenes: M = 3.08, SD = 1.13), $t(23) = -0.21$, p = .83, d = –0.04, or among all blocks (congruent scenes: M = 2.34, SD = 0.77; incongruent scenes: M = 2.36, SD = 0.77), $t(23) = -0.69$, p = .50, d = –0.14. Similarly, a BF analysis of the data always indicated convincing evidence for the absence of a scene congruency effect (BF = 17 and BF = 32, for the first block and all data, respectively).

As can be derived from the figure depicting the results of the sequential analysis, we crossed the significance boundary twice during the data collection process when only the first block would be considered. Increasing the number of trials per observer, however, never led to a significant main effect of congruency. Again, and in contrast to the results of the traditional analysis, the BF analysis showed a gradual increase in evidence for the absence of a scene congruency effect as the data came in.

Supplementary analyses indicated that the results did not change when participants' categorization performance on the post-experimental rating session was taken into account nor when only participants were included for whom the CFS mask was presented to the dominant eye.

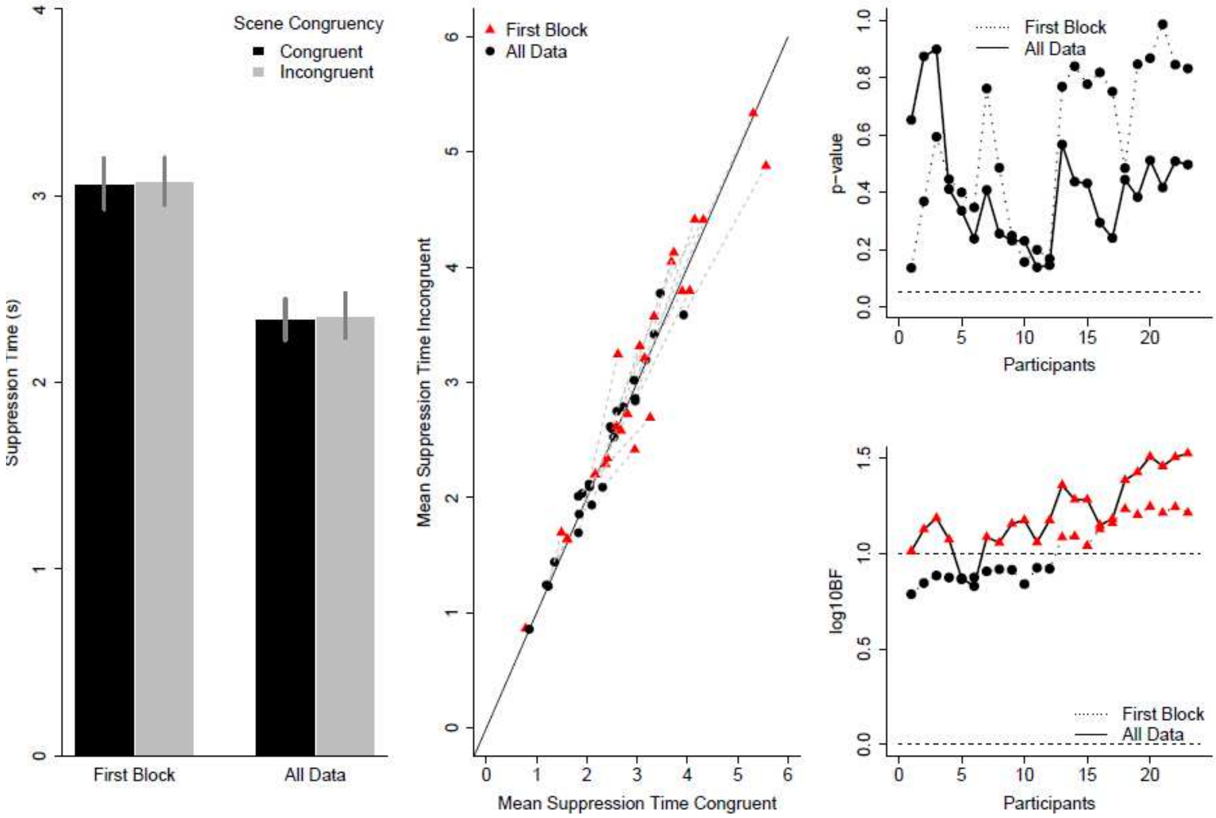

Figure 4. Results of Experiment 2. (left) Bar plot depicting mean suppression times for the first block only and all data. Error bars refer to 95% within-subject confidence intervals with the adjustment suggested by Morey (2008). (middle) Scatter plot of mean suppression times for congruent and incongruent scenes for each participant. Red triangles refer to the data of the first block, black dots to all data together. The gray dashed lines connect the same participants. (right) Sequential analysis of the data. (top row) Evolution of p-values as more data were collected for the effect of scene congruency. Black symbols indicate p-values higher than .05 whereas red symbols indicate p-values smaller than .05. (lower row) Evolution of Bayes Factors. Bayes Factors higher than 0 indicate evidence for the absence of an effect whereas Bayes Factors smaller than 0 indicate evidence for the presence of an effect.

DISCUSSION

Although we did not obtain any evidence in favor of a scene congruency effect in both our experiments, one might still raise objections to our current attempts. That is, compared to the original study, we did not implement a post-experimental rating procedure in which participants were asked to categorize the images as being unusual or not (i.e., incongruent versus congruent). Second, the CFS mask was not presented in every participants' dominant eye. Third, the addition of an inversion condition in Experiment 1 and dropping the mask fade-out procedure as well as repeating the images more than once in Experiment 2 might have obscured a subtle congruency effect. We addressed these concerns in a third experiment, in which we presented every scene upright and only once across two experimental blocks and always in the participants' non-dominant eye. We included the mask fade-out procedure again, and implemented a post-experimental rating session. At this point, we invited participants from Experiments 1 and 2 to conduct this rating session on-line and reanalyzed the data based on their responses.

EXPERIMENT 3

METHODS

Participants

A new sample of 50 people participated in the experiment in return for money or course credit. All participants were naïve with respect to the design of the study and had normal or corrected-to-normal vision. All participants provided written informed consent prior to the start of the experiment.

Apparatus and Stimuli

The experimental set-up and stimuli were exactly the same as in Experiments 1 and 2.

Procedure and Design

Up to the post-experimental rating session, the procedure was exactly the same as in Experiments 1 and 2. Participants first completed the eye dominance experiment after which they completed two blocks of trials in which the scenes were presented to the participants non-dominant eye and the CFS mask to the dominant eye, according to the same procedure as in the previous experiments. After completing the CFS experiment, each scene was presented binocularly and participants had to indicate whether they thought the presented scene was unusual our not.

The experiment consisted of a within-subject design with two conditions (i.e., congruent versus incongruent scenes). In the main experiment, participants completed 8 practice trials, and 80 experimental trials. Within a block, the order in which trials were presented was completely randomized. The post-experimental rating session also consisted of 80 trials, presented in a random order.

RESULTS

The data of Experiment 3 were analyzed in the same way as in Experiments 1 and 2. That is, analyses were performed on correct trials (M =0.98, SD = 0.03) including only trials of which the suppression time did not exceed the time at which the mask reached 0% contrast (M = 0.91, SD = 0.18), after removing outliers defined as suppression times higher than the mean suppression time plus three times the standard deviation (for observers separately, M = 0.005, SD = 0.01). Furthermore, only stimuli that were correctly categorized after the experiment (i.e., congruent as incongruent and vice versa) were included in the analysis (M = 0.71, SD = 0.06). An analysis including the incorrectly rated scenes (see Supplementary Materials) did not change the results of the analysis.

The consistency of which stimuli broke suppression fast and slow (across observers) was similar to what was observed in Experiments 1 and 2 (α = 0.85). Furthermore, the ordering again correlated very strongly with the ordering observed in Experiment 1 (r = .89, BF < 0.01) and Experiment 2 (r = .94, BF < 0.01), indicating that the images (irrespective of their congruency) were processed similarly across all three experiments.

As in Experiments 1 and 2, an analysis of mean correct suppression times revealed no effects of scene congruency (congruent scenes: M = 2.65, SD = 1.14; incongruent scenes: M = 2.65, SD = 1.09), $t(49) = -0.03$, p = .97, $d = -0.005$. Similarly, the BF analysis indicated strong evidence for the absence of a congruency effect (BF = 21). Figure 5 depicts the results of the sequential analysis. As is apparent from this figure, we observed a significant congruency effect in the classical analysis early on during data collection, yet the BF never crossed the boundary for indicating evidence in favor of a congruency effect and again gradually accumulated evidence in favor of the absence of a congruency effect.

In a supplementary exploratory analysis (see Supplementary Materials), we examined the relationship between various statistical image properties and the mean suppression times of the images in all three experiments. That is, given the absence of a scene congruency effect, yet high consistency of which images broke suppression slow and fast, we were interested in exploring whether some statistical properties of the images would predict suppression times. The results indicated that a measure of spatial coherence (an indicator of scene fragmentation) correlated positively with mean suppression duration. That is, when images were more cluttered, suppression times were on average higher.

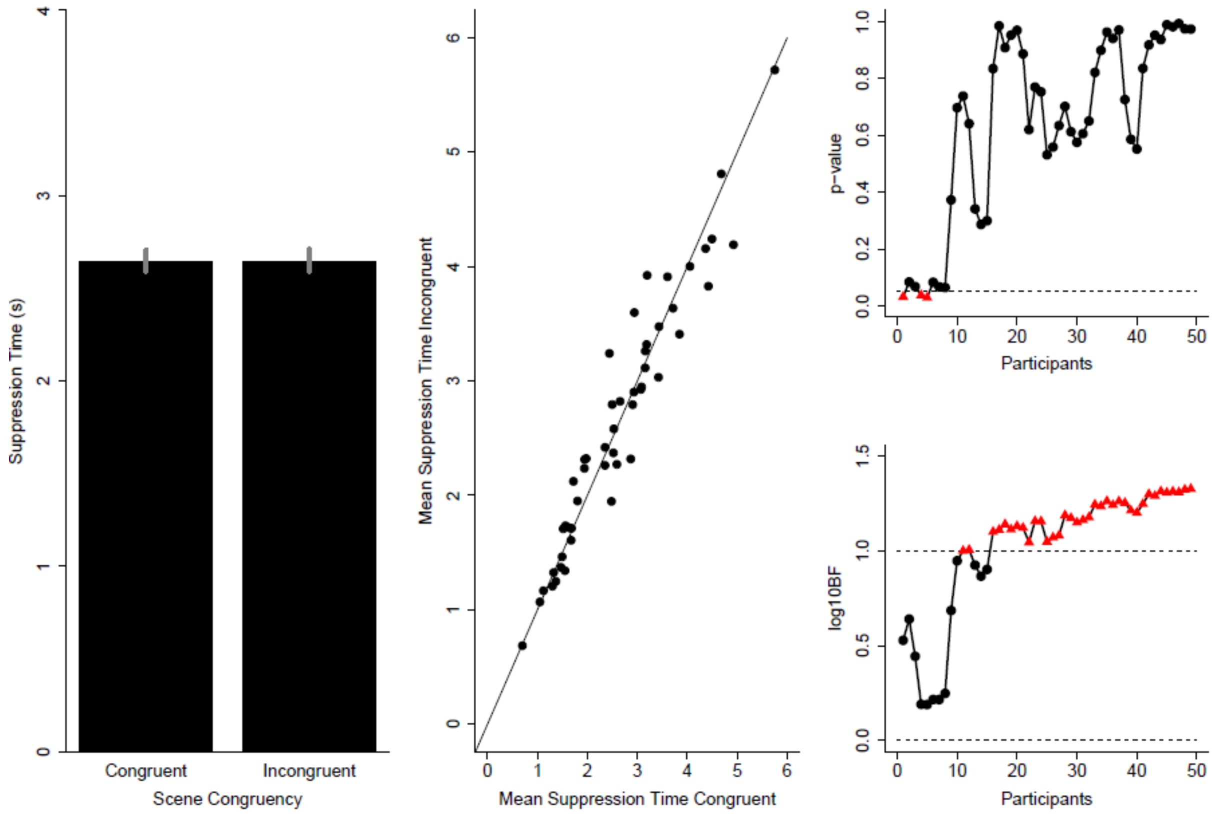

Figure 5. Results of Experiment 3. (left) Bar plot depicting mean suppression times for both conditions. Error bars refer to 95% within-subject confidence intervals with the adjustment suggested by Morey (2008). (middle) Scatter plot of mean suppression times for congruent and incongruent scenes for each participant. (right) Sequential analysis of the data. (top row) Evolution of p-values as more data were collected for the effect of scene congruency. Black symbols indicate p-values higher than .05 whereas red symbols indicate p-values smaller than .05. (lower row) Evolution of Bayes Factors. Log10BFs higher than 0 indicate evidence for the absence of a congruency effect whereas log10BFs smaller than 0 indicate evidence for the presence of a congruency effect.

DISCUSSION AND CONCLUSION

The goal of this study was to assess the replicability of the findings reported in Mudrik et al. (2011) and to gauge whether the scene congruency effect was attributable to low-level differences between the stimuli by including a scene inversion condition. In Experiment 1, using a larger sample as in the original study, we did not observe an effect of scene congruency, and critically, no interaction between scene inversion and scene congruency, highlighting that the inversion did not modulate the effect of scene congruency, which would have been expected if the scene congruency effect was a semantic effect. Furthermore, a Bayes Factor (BF) analysis relying on linear mixed-effects models with crossed random effects showed convincing evidence for the absence of a scene congruency effect. A sequential analysis of our data highlighted a strong discrepancy between the inferences based on a classical repeated measures ANOVA and the BF analysis. Indeed, at several steps throughout our data collection, the classical analysis yielded a significant main effect of scene congruency whereas this was never the case for the BF analysis, indicating that not considering the random item variation yields a too liberal statistical procedure (Clark, 1973). Interestingly, in the case of the scene inversion effect, both the classical and the Bayesian analysis converged to the same conclusion. In Experiment 2, we increased the number of trials fourfold for each observer to assess whether increasing the precision of the effect size estimate for each observer would reveal a more consistent scene congruency effect. Again, we did not observe a scene congruency effect. Experiment 3 consisted of a third high-powered replication attempt in which we also included a post-experimental rating task to exclude stimuli that were not perceived as (in)congruent by our participants. The results of this last experiment also indicated strong evidence for the absence of a scene congruency effect.

What do the results of these experiments tell us? Could it be that the images were not processed at all? We contend that several aspects of our results argue against such an interpretation. In all experiments, we observed high internal consistency in which items broke suppression fast and slow across observers. Moreover, this pattern of slow and fast items correlated strongly across all experiments in independent sets of observers. Third, a measure of spatial coherence correlated with suppression times (with a similar magnitude) in all experiments. These aspects indicate that the scene stimuli did not break suppression in a completely random fashion.

In Experiment 1, we observed a consistent scene inversion effect. Could this be an indication that the stimuli were somehow interpreted during suppression? Although inversion reduces the identifiability of the scenes and preserves low-level image statistics, it may also influence higher-order image statistics to which the visual system is sensitive (Okazawa, Tajima, & Komatsu, 2015). Therefore, an inversion effect does not by definition indicate that a stimulus is processed up to a semantic level, yet it could also reflect the sensitivity of the visual system to natural input.

Although the original scene congruency effect was interpreted as evidence for unconscious integration, our results cannot be interpreted as providing evidence *against* unconscious integration per se. That is, there remain some differences between our study and the original one. Besides the obvious differences in hardware, experimental environment, and pools of observers, there were also slight differences in our trial randomization procedure and the size of our CFS mask display. Our findings therefore show that the results of Mudrik et al. (2011) do not generalize across these particular testing differences, indicating that, if true, the scene congruency effect is particularly fragile. However, this lack of generalizability is hardly compatible with the conclusions derived from the original result. If unconscious integration can manifest under CFS, such an effect should not be dependent upon factors such as particular testing conditions or a different participant pool. Indeed, what our results show is that there is no evidence for scene integration without awareness under CFS. Indeed, although CFS initially proved to be a promising technique to assess the limits of unconscious visual processing (Bahrami et al., 2010; Jiang, Costello, Fang, Huang, & He, 2006; Jiang, Costello, & He, 2007; Sklar et al., 2012), our findings fit well in a series of more recent findings highlighting rather limited visual processing during CFS (Hedger, Adams, & Garner, 2015; Hesselmann & Knops, 2014; Heyman & Moors, 2014; Moors, Huygelier, Wagemans, de-Wit, & van Ee, 2015; Moors, Wagemans, van Ee, & de-Wit, 2015). In hindsight, this is also not too surprising given that it is known that binocular rivalry disrupts processing of the suppressed stimulus beyond early visual areas (Fogelson, Kohler, Miller, Granger, & Tse, 2014; Hesselmann & Malach, 2011; Yuval-Greenberg & Heeger, 2013). This does not imply, however, that unconscious integration per se cannot take place. Indeed, some forms of unconscious integration have been shown to exist (see Mudrik, Faivre, & Koch, 2014 for a review), yet often relying on different suppression paradigms.

While this study highlights the importance of replication and the inclusion of appropriate control conditions, it also reveals a much broader point. That is, it provides an important example of how different statistical methods can strongly disagree throughout the data collection process. Indeed, although both types of analysis yielded the same conclusion at the end of data collection, it is important to highlight that the traditional repeated-measures ANOVA more than once indicated a significant scene congruency effect as the data came in. The BF analysis, however, yielded a more consistent picture in that it always provided evidence for the absence of a scene congruency effect while the evidence also gradually accumulated when more data was collected. Furthermore, in Experiment 1, the BF indicated convincing evidence for the presence of an inversion effect and here, the results of the repeated measures ANOVA converged on those of the BF analysis. This highlights that for experimental designs in which the dependent measure can vary across participants and items, the classical repeated measures ANOVA approach might be too liberal (Clark, 1973) and an approach based on crossed random effects is recommended (Baayen, Davidson, & Bates, 2008).

In sum, our study questions the replicability and generalizability of the findings reported in Mudrik et al. (2011) by obtaining strong evidence for the absence of a scene congruency effect across three experiments and moreover showing that scene congruency was not modulated by scene inversion. Therefore, it is unlikely that during CFS, complex high-level scene processing can ensue.

AUTHOR CONTRIBUTIONS

P. Moors developed the study concept. Testing and data collection were performed by D. Boelens and J. van Overwalle. P. Moors performed the data analysis and interpretation. P. Moors and J. Wagemans drafted the manuscript and D. Boelens and J. van Overwalle provided critical revisions. All authors approved the final version of the manuscript for submission.

ACKNOWLEDGMENTS

This research was supported by a doctoral fellowship awarded to P. Moors by the Fund for Scientific Research Flanders (FWO) and by Methusalem program of the Flemish government (METH/08/02 and METH/14/02) awarded to J. Wagemans. We are grateful to Liad Mudrik for sharing the original stimulus set.

REFERENCES

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Bahrami, B., Vetter, P., Spolaore, E., Pagano, S., Butterworth, B., & Rees, G. (2010). Unconscious numerical priming despite interocular suppression. *Psychological Science*, *21*(2), 224–233. http://doi.org/10.1177/0956797609360664

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. http://doi.org/10.1016/S0022-5371(73)80014-3

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227. http://doi.org/10.1016/j.neuron.2011.03.018

Fogelson, S. V., Kohler, P. J., Miller, K. J., Granger, R., & Tse, P. U. (2014). Unconscious neural processing differs with method used to render stimuli invisible. *Frontiers in Psychology*, *5*, 601. http://doi.org/10.3389/fpsyg.2014.00601

Hedger, N., Adams, W. J., & Garner, M. (2015). Autonomic arousal and attentional orienting to visual threat are predicted by awareness. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 798–806. http://doi.org/10.1037/xhp0000051

Hesselmann, G., & Knops, A. (2014). No conclusive evidence for numerical priming under interocular suppression. *Psychological Science*. http://doi.org/10.1177/0956797614548876

Hesselmann, G., & Malach, R. (2011). The link between fMRI-BOLD activation and perceptual awareness is “stream-invariant” in the human visual system. *Cerebral Cortex*, *21*(12), 2829 – 2837. http://doi.org/10.1093/cercor/bhr085

Heyman, T., & Moors, P. (2014). Frequent words do not break continuous flash suppression differently from infrequent or nonexistent words: Implications for semantic processing of words in the absence of awareness. *PLoS ONE*, *9*(8), e104719. http://doi.org/10.1371/journal.pone.0104719

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506.

Jeffreys, H. (1961). *Theory of Probability*. Oxford, UK: Oxford University Press.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(45), 17048–17052. http://doi.org/10.1073/pnas.0605678103

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science*, *18*(4), 349–355. http://doi.org/10.1111/j.1467-9280.2007.01902.x

Moors, P., Huygelier, H., Wagemans, J., de-Wit, L., & van Ee, R. (2015). Suppressed visual looming stimuli are not integrated with auditory looming signals: Evidence from continuous flash suppression. *I-Perception*, *6*(1), 48–62. http://doi.org/doi:10.1068/i0678

Moors, P., Wagemans, J., van Ee, R., & de-Wit, L. (2015). No evidence for surface organization in Kanizsa configurations during continuous flash suppression. *Attention, Perception & Psychophysics*, Ahead of print. http://doi.org/10.3758/s13414-015-1043-x

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: 0.9.11-1 CRAN* [Zenodo]. Retrieved from http://doi.org/10.5281/zenodo.16238

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, *22*(6), 764–770. http://doi.org/10.1177/0956797611408736

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496. http://doi.org/10.1016/j.tics.2014.04.009

Neumann, D., & Gegenfurtner, K. R. (2006). Image retrieval and perceptual similarity. *ACM Transactions on Applied Perception (TAP)*, *3*(1), 31–47. http://doi.org/10.1145/1119766.1119769

Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proceedings of the National Academy of Sciences*, *112*(4), E351–E360. http://doi.org/10.1073/pnas.1415146112

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*(10). http://doi.org/10.3389/neuro.11.010.2008

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. http://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. http://doi.org/10.3758/PBR.16.2.225

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences*, *109*(48), 19614–19619. http://doi.org/10.1073/pnas.1211645109

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. http://doi.org/10.1038/nn1500

Yang, E., Blake, R., & McDonald, J. E. (2010). A new interocular suppression technique for measuring sensory eye dominance. *Investigative Ophthalmology & Visual Science*, *51*(1), 588–593. http://doi.org/10.1167/iovs.08-3076

Yuval-Greenberg, S., & Heeger, D. J. (2013). Continuous flash suppression modulates cortical activity in early visual cortex. *The Journal of Neuroscience*, *33*(23), 9635–9643. http://doi.org/10.1523/JNEUROSCI.4612-12.2013

Supplementary Materials

Scene integration without awareness: No conclusive evidence for processing scene congruency during continuous flash suppression.

Pieter Moors, David Boelens, Jaana van Overwalle, & Johan Wagemans

In this Supplementary Material, we report on additional analyses for Experiments 1 and 2 as well as an exploratory analysis on the relationship between mean suppression times obtained for each image averaged across observers and four statistical properties of the images.

In Experiment 3, we implemented a post-experimental rating procedure in which participants had to indicate whether they thought a presented scene was unusual or not. This procedure was not implemented in Experiments 1 and 2 and might have influenced the results. Therefore, we invited the participants of the first two experiments to participate in an on-line experiment in which they had to rate all 80 scenes. The data of the first two experiments were then reanalyzed by first excluding incorrectly rated scenes (i.e., an incongruent scene was rated as congruent and vice versa). Second, in Experiment 3, a technical issue was also fixed. That is, due to a bug in the code, after the eye dominance measurement, the CFS mask was always presented to the right eye rather than the dominant eye. For Experiments 1 and 2, we also reanalyzed the data by excluding participants for which the CFS mask was not presented to the dominant eye (i.e., participants for whom the dominant eye was the left one).

Supplementary analysis for Experiment 1

23 out of 45 (51%) participants responded to our invitation to participate in the rating experiment. Figures S1 and S2 depict the results of Experiment 1 for these 23 participants after also excluding all data points from incorrectly rated scenes. The results are very similar to those observed when analyzing the full data set. That is, a two-way repeated measures ANOVA on the mean correct suppression times revealed no main effect of scene congruency ($M_{congruent}$ = 2.36, $M_{incongruent}$ = 2.35; $SD_{congruent}$ = 0.94, $SD_{incongruent}$ = 0.88; $F(1,22) = 0.095$, $p = .76$, $d = 0.05$), a main effect of scene inversion ($M_{upright}$ = 2.31, $M_{inverted}$ = 2.40; $SD_{upright}$ = 0.92, $SD_{inverted}$ = 0.90; $F(1,22) = 5.43$, $p = .03$, $d = -0.37$), and no interaction between scene congruency and scene inversion ($F(1,22) = 0.068$, $p = .78$, $d = -0.05$). The Bayes Factor analysis indicated strong evidence in favor of the absence of a scene congruency effect (BF = 22), moderate evidence in favor of the presence of a scene inversion effect (BF = 6) and strong evidence for the absence of an interaction effect (BF = 16).

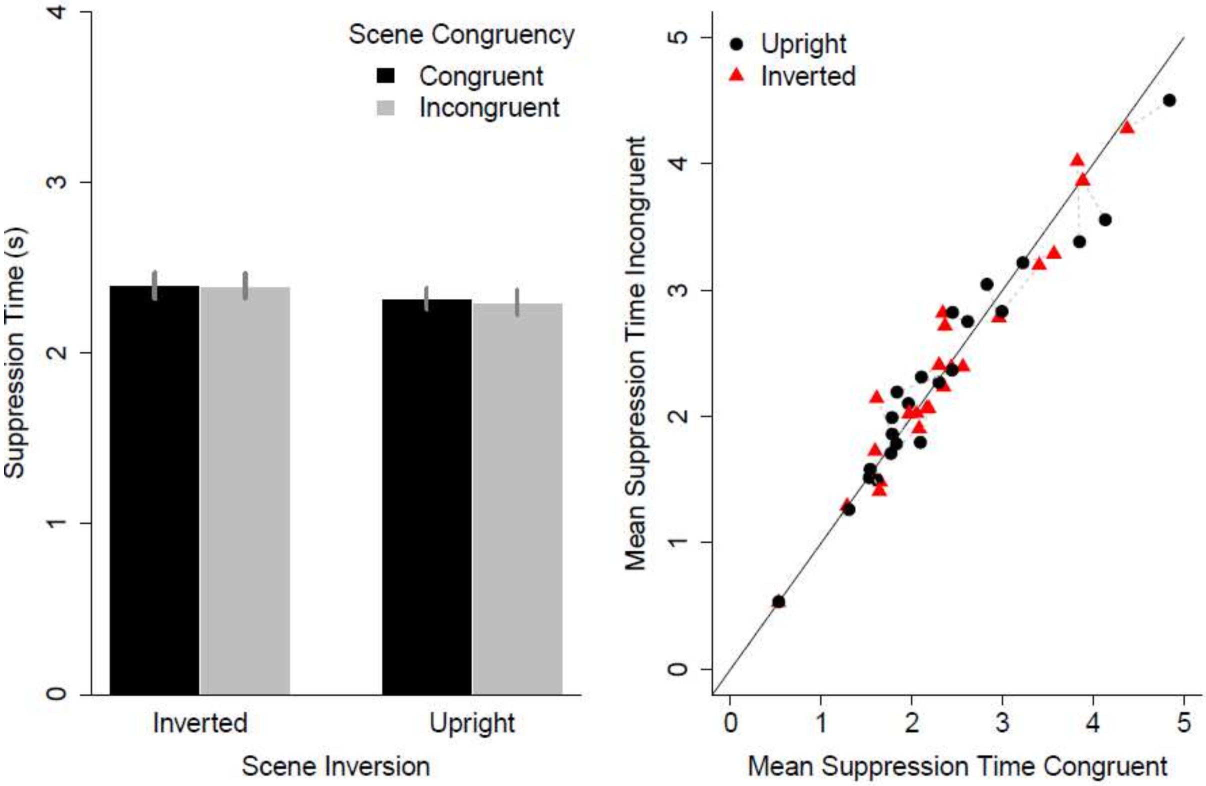

Figure S1. Results of Experiment 1 after removing the incorrectly categorized scenes. Mean suppression time (a) is graphed as a function of scene inversion, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown separately for the two scene types. Each dashed line connects the data points for upright and inverted scenes for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes.

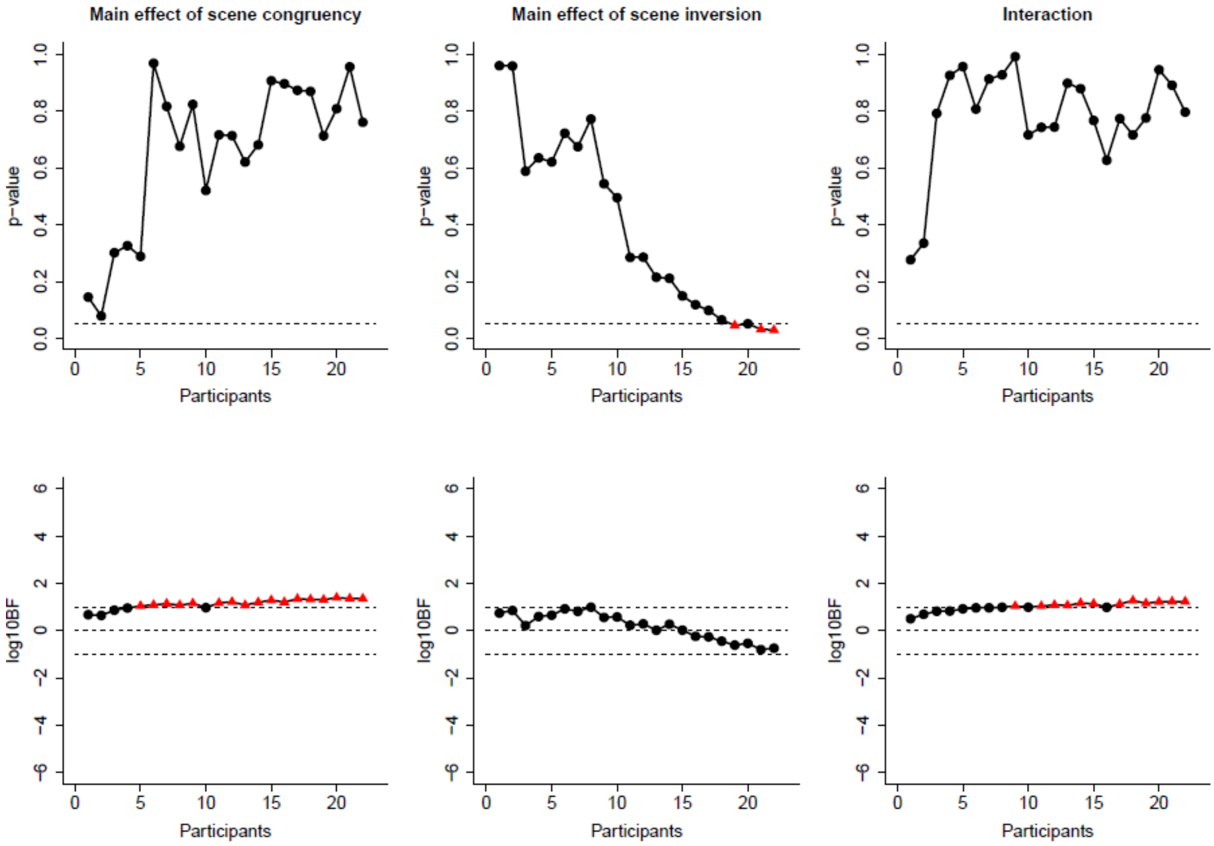

Figure S2. Sequential analysis of the data from Experiment 1 after removing incorrectly categorized scenes. The graphs show *p* values (top row) and Bayes factor values (bottom row) as a function of the number of participants included in the analysis, separately for the main effect of scene congruency, the main effect of scene inversion, and their interaction. Bayes factors (log10) greater than 0 indicate evidence for the absence of an effect, whereas Bayes factors smaller than 0 indicate evidence for the presence of an effect.

We conducted the same analysis, but now excluding all participants for which the CFS mask was not presented to the dominant eye. This reduced our sample to 25 participants (20 were excluded). The results of this analysis are depicted in Figures S3 and S4. Again, the results are very similar to those observed when analyzing the full data set. That is, a two-way repeated measures ANOVA on the mean correct suppression times revealed no main effect of scene congruency ($M_{congruent}$ = 2.73, $M_{incongruent}$ = 2.67; $SD_{congruent}$ = 1, $SD_{incongruent}$ = 0.97; $F(1,24) = 2.94$, $p = .1$, $d = 0.21$), no main effect of scene inversion ($M_{upright}$ = 2.67, $M_{inverted}$ = 2.73; $SD_{upright}$ = 1.01, $SD_{inverted}$ = 0.96; $F(1,24) = 2.13$, $p = .16$, $d = -0.20$), and no interaction between scene congruency and scene inversion ($F(1,24) = 0.05$, $p = .82$, $d = -0.05$). The Bayes Factor analysis indicated evidence in favor of the absence of a scene congruency effect (BF = 6), evidence in favor of the presence of a scene inversion effect (BF = 4) and strong evidence for the absence of an interaction effect (BF = 17).

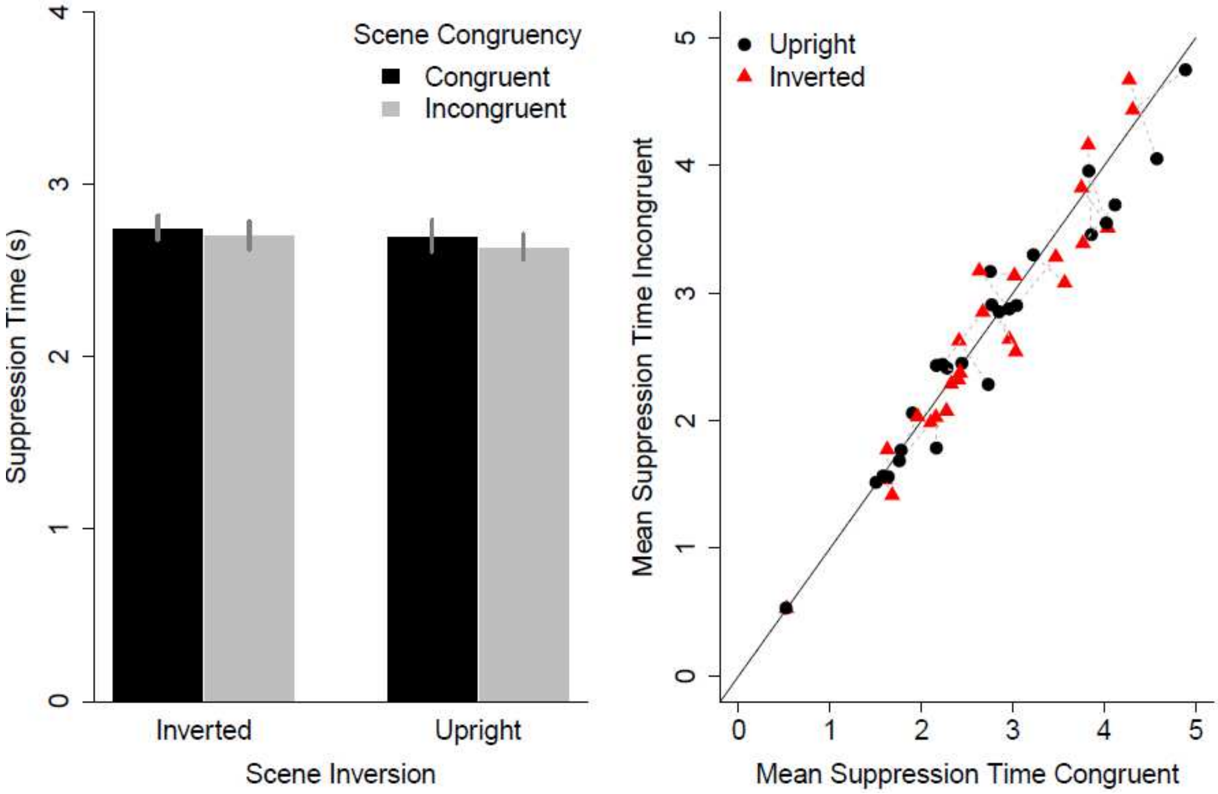

Figure S3. Results of Experiment 1 after removing participants from whom the CFS mask was presented in the nondominant eye. Mean suppression time (a) is graphed as a function of scene inversion, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown separately for the two scene types. Each dashed line connects the data points for upright and inverted scenes for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes.

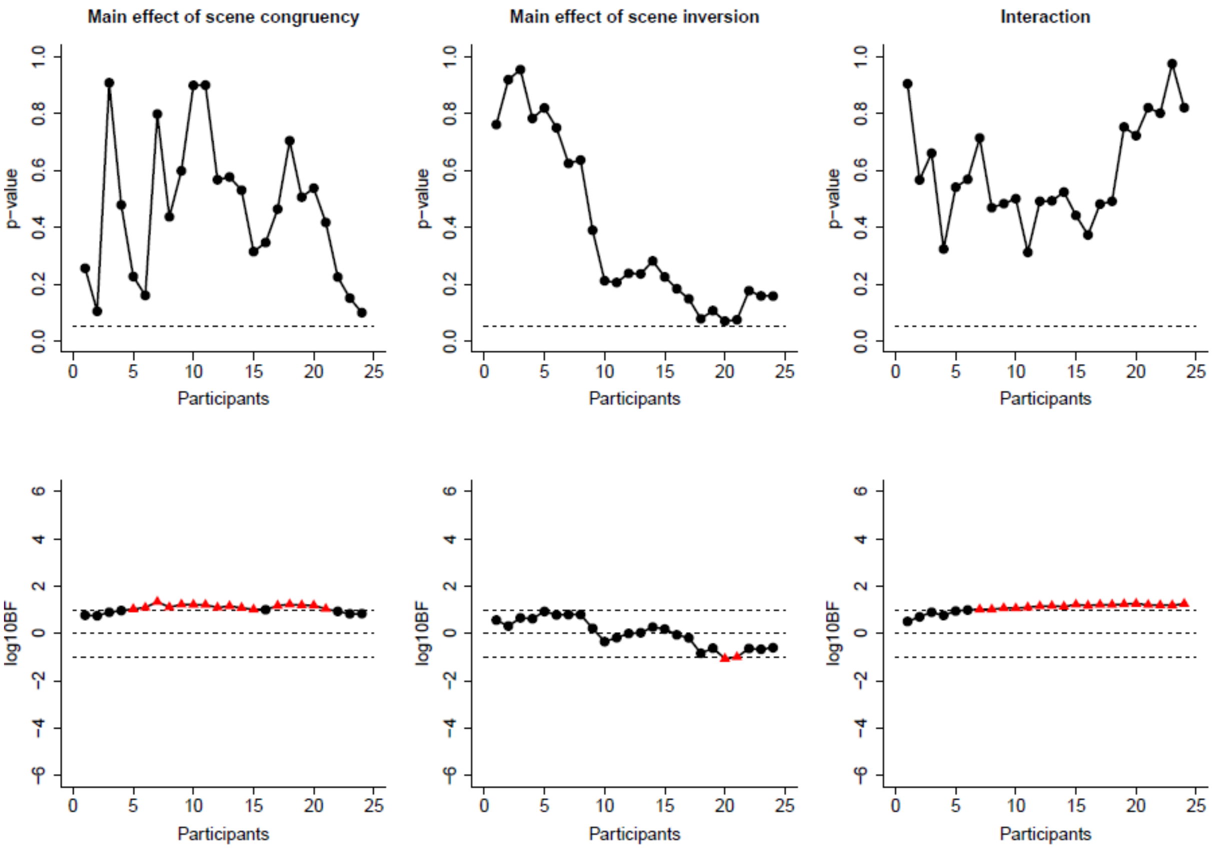

Figure S4. Sequential analysis of the data from Experiment 1 after removing participants for whom the CFS mask was presented in the nondominant eye. The graphs show *p* values (top row) and Bayes factor values (bottom row) as a function of the number of participants included in the analysis, separately for the main effect of scene congruency, the main effect of scene inversion, and their interaction. Bayes factors (log10) greater than 0 indicate evidence for the absence of an effect, whereas Bayes factors smaller than 0 indicate evidence for the presence of an effect.

For the sake of completeness, including only those people that responded to the invitation to participate in the rating experiment and those for which the CFS mask was presented to the dominant eye further reduced our sample to 15 participants. This analysis yielded slightly different results compared to the analysis of the full data set and is depicted in Figures S5 and S6. That is, a two-way repeated measures ANOVA on the mean correct suppression times revealed a main effect of scene congruency ($M_{congruent}$ = 2.53, $M_{incongruent}$ = 2.43; $SD_{congruent}$ = 1.09, $SD_{incongruent}$ = 1.04; $F(1,14) = 18.82$, $p = .0007$, $d = 0.61$), no main effect of scene inversion ($M_{upright}$ = 2.44, $M_{inverted}$ = 2.52; $SD_{upright}$ = 1.08, $SD_{inverted}$ = 1.05; $F(1,14) = 3.29$, $p = .09$, $d = -0.37$), and, critically, no interaction between scene congruency and scene inversion (F(1,14) = 0.33, p = .57, $d = -0.15$). The Bayes Factor analysis did not indicate a preference for the presence or absence of a congruency effect (BF = 1.3), nor evidence in favor of the presence or absence of a scene inversion effect (BF = 0.54) and indicated strong evidence for the absence of an interaction effect (BF = 13).

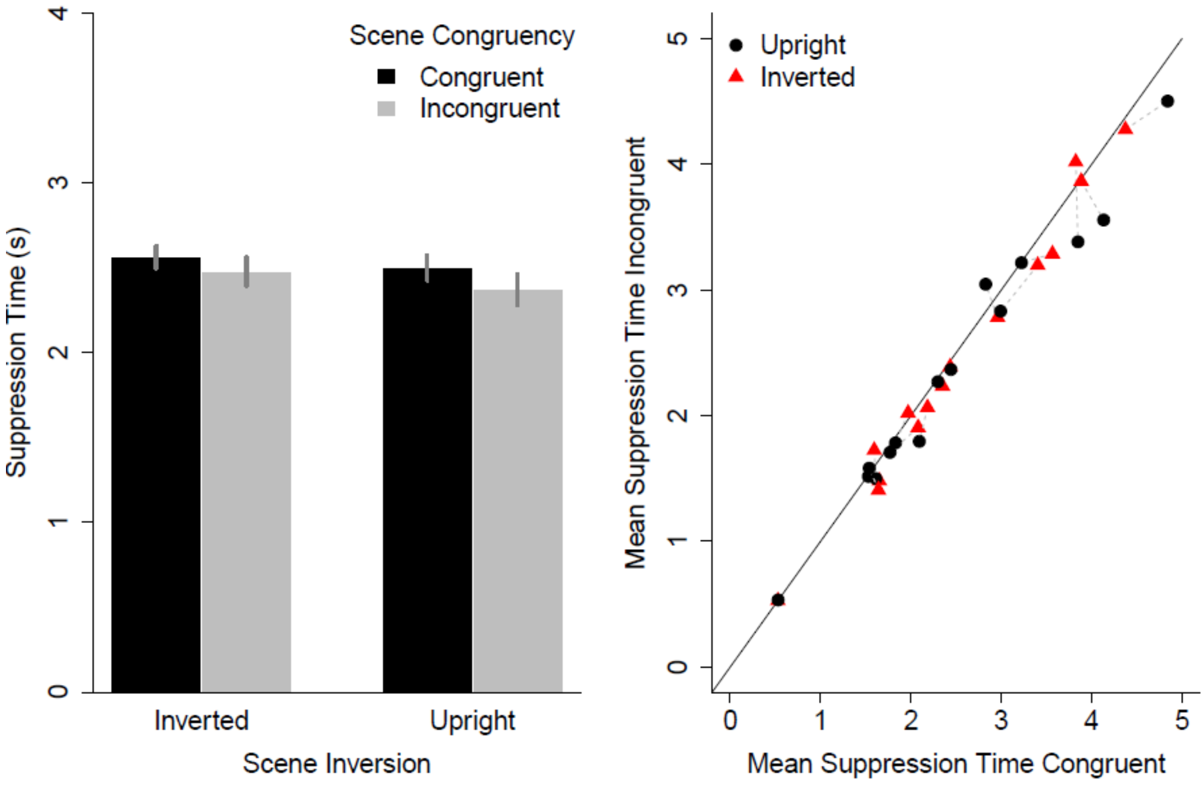

Figure S5. Results of Experiment 1 after removing incorrectly categorized scenes and participants for whom the CFS mask was presented in the nondominant eye. Mean suppression time (a) is graphed as a function of scene inversion, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown separately for the two scene types. Each dashed line connects the data points for upright and inverted scenes for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes.

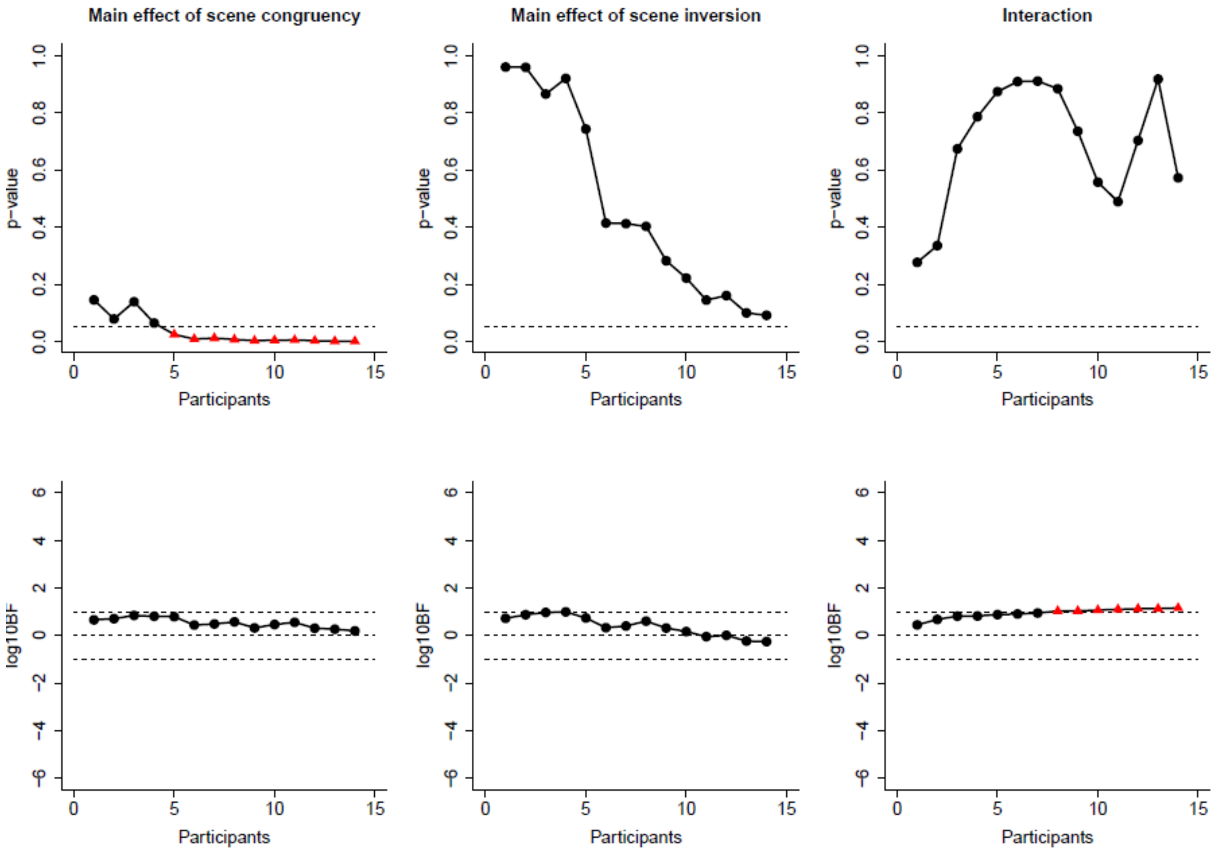

Figure S6. Sequential analysis of the data from Experiment 1. The graphs show *p* values (top row) and Bayes factor values (bottom row) as a function of the number of participants included in the analysis, separately for the main effect of scene congruency, the main effect of scene inversion, and their interaction. Bayes factors (log10) greater than 0 indicate evidence for the absence of an effect, whereas Bayes factors smaller than 0 indicate evidence for the presence of an effect.

Supplementary analysis for Experiment 2

We now report on the same reanalysis of the data of Experiment 2 by first considering those participants that performed the rating experiment, then considering those for which the CFS mask was presented in the dominant eye, and ending with combining both exclusion criteria.

18 out of 24 (75%) participants responded to our invitation to participate in the rating experiment. Excluding all incorrectly rated scenes yielded a similar picture as the results for the full data set. That is, no effect of scene congruency was observed neither in the first block ($M_{congruent}$ = 2.92, $M_{incongruent}$ = 2.95; $SD_{congruent}$ = 1.12, $SD_{incongruent}$ = 1.07; $t(17) = -0.36$, $p = .72$, $d = -0.09$) nor when considering the data as a whole ($M_{congruent}$ = 2.23, $M_{incongruent}$ = 2.27; $SD_{congruent}$ = 0.71, $SD_{incongruent}$ = 0.70; $t(17) = -1.37$, $p = .19$, $d = -0.32$). Similarly, a BF analysis of the data always indicated convincing evidence for the absence of a scene congruency effect (BF = 10 and BF = 15, for the first block and all data, respectively). Figure S7 depicts the results of this analysis.

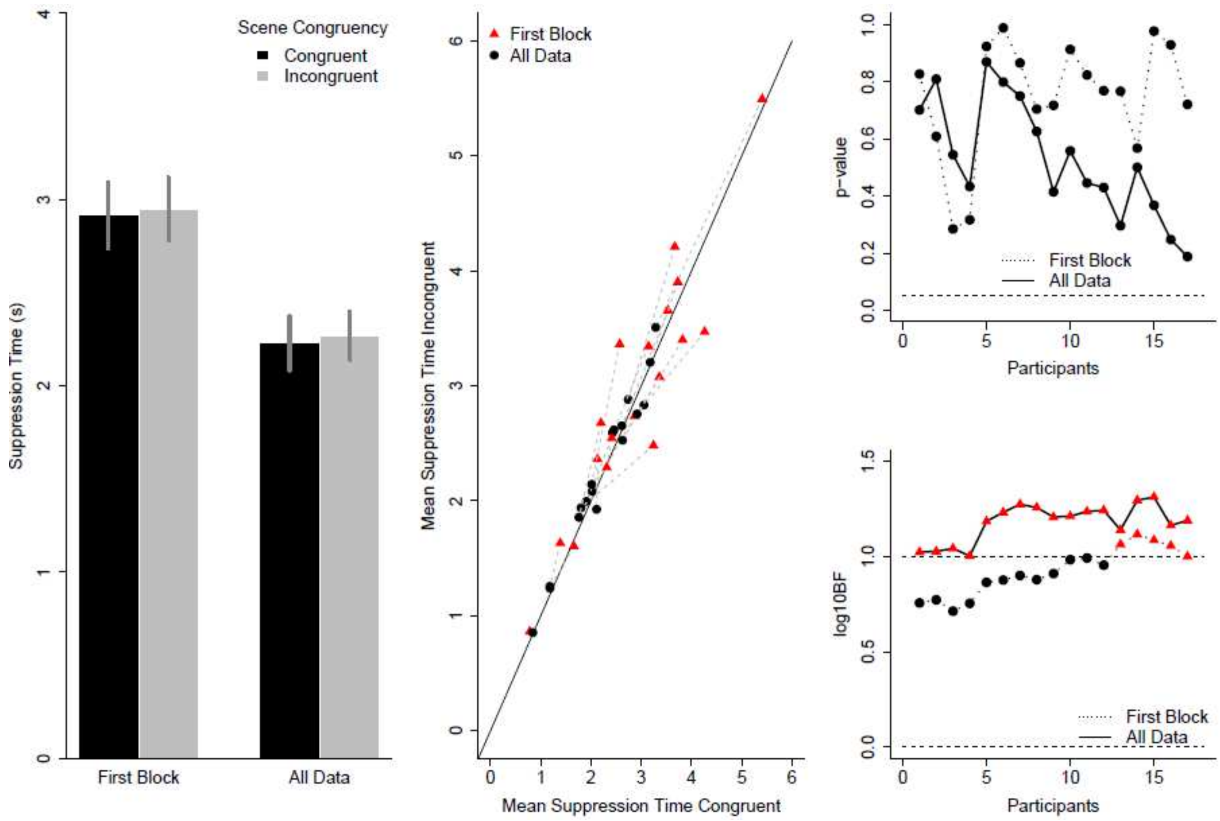

Figure S7. Results of Experiment 2 after removing incorrectly categorized scenes. Mean suppression time (a) is graphed for the first block and for all blocks, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown for each participant's performance in the first block and in all blocks. Each dashed line connects the data points for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes. The two other graphs show *p* values (c) and Bayes factor values (d) as a function of the number of participants in the analysis, separately for the first block and for all blocks. Bayes factors (log10) greater than 0 indicate evidence for the absence of an effect, whereas Bayes factors smaller than 0 indicate evidence for the presence of an effect.

Including only those participants for which the CFS mask was presented to the dominant eye reduced our sample to 12 observers. The results are summarized in Figure S8 and are very similar to the analyses for the full data set. That is, no effect of scene congruency was observed neither in the first block ($M_{congruent}$ = 3.75, $M_{incongruent}$ = 3.71; $SD_{congruent}$ = 0.98, $SD_{incongruent}$ = 0.93; $t(11)$ = 0.37, p = .72, d = 0.11) nor when considering the data as a whole ($M_{congruent}$ = 2.77, $M_{incongruent}$ = 2.80; $SD_{congruent}$ = 0.66, $SD_{incongruent}$ = 0.65; $t(11)$ = -0.53, p = .61, d = -0.15). Similarly, a BF analysis of the data always indicated convincing evidence for the absence of a scene congruency effect (BF = 13 and BF = 26, for the first block and all data, respectively).

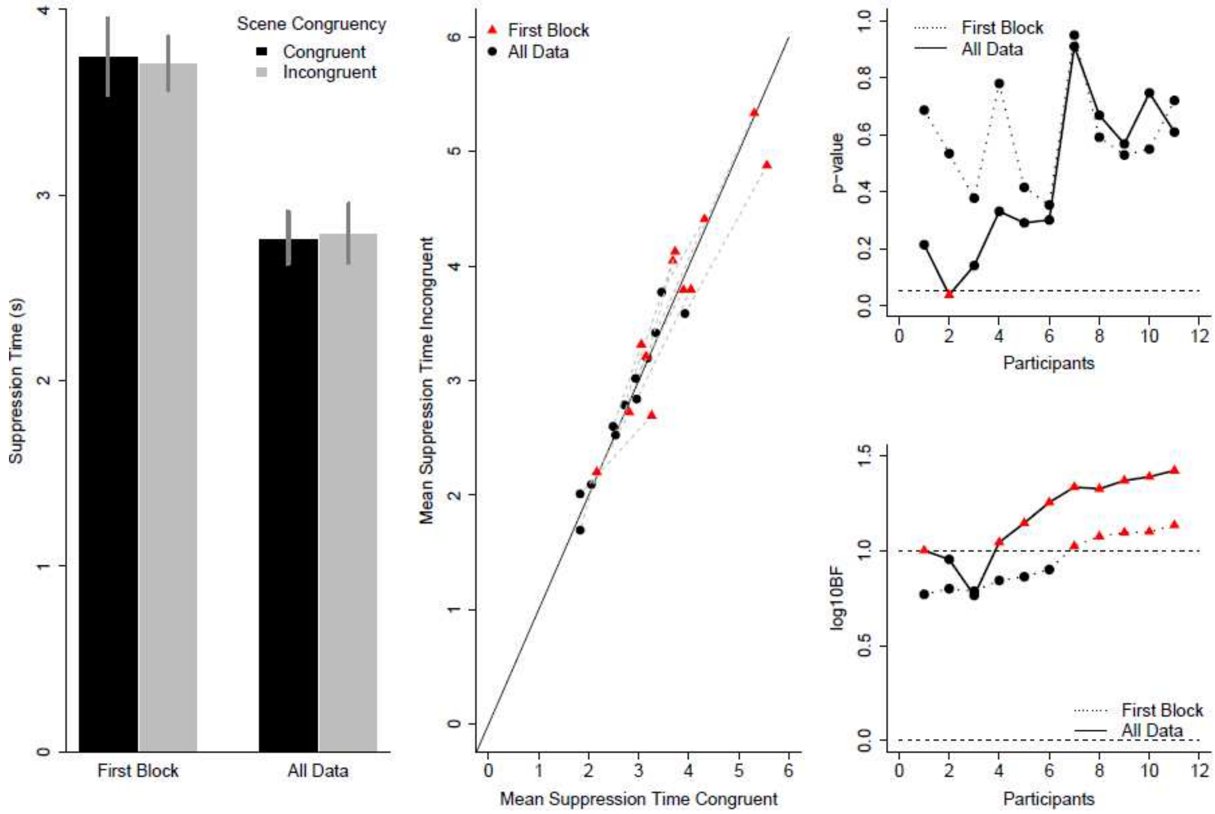

Figure S8. Results of Experiment 2 after removing participants for whom the CFS mask was presented in the nondominant eye. Mean suppression time (a) is graphed for the first block and for all blocks, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown for each participant's performance in the first block and in all blocks. Each dashed line connects the data points for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes. The two other graphs show *p* values (c) and Bayes factor values (d) as a function of the number of participants in the analysis, separately for the first block and for all blocks. Bayes factors (log10) greater than 0 indicate evidence for the absence of an effect, whereas Bayes factors smaller than 0 indicate evidence for the presence of an effect.

Last, an analysis combining both exclusion criteria further reduced our sample to 9 participants. The results of this analysis are summarized in Figure S9. Again, these results are very similar to those observed for the full data set. No effect of scene congruency was observed neither in the first block ($M_{congruent}$ = 3.72, $M_{incongruent}$ = 3.57; $SD_{congruent}$ = 0.75, $SD_{incongruent}$ = 0.90; $t(8) = 1.02$, $p = .34$, $d = 0.34$) nor when considering the data as a whole ($M_{congruent}$ = 2.68, $M_{incongruent}$ = 2.71; $SD_{congruent}$ = 0.51, $SD_{incongruent}$ = 0.49; $t(8) = -0.57$, $p = .59$, $d = -0.19$). Similarly, a BF analysis of the data always indicated convincing evidence for the absence of a scene congruency effect (BF = 10 and BF = 13, for the first block and all data, respectively).

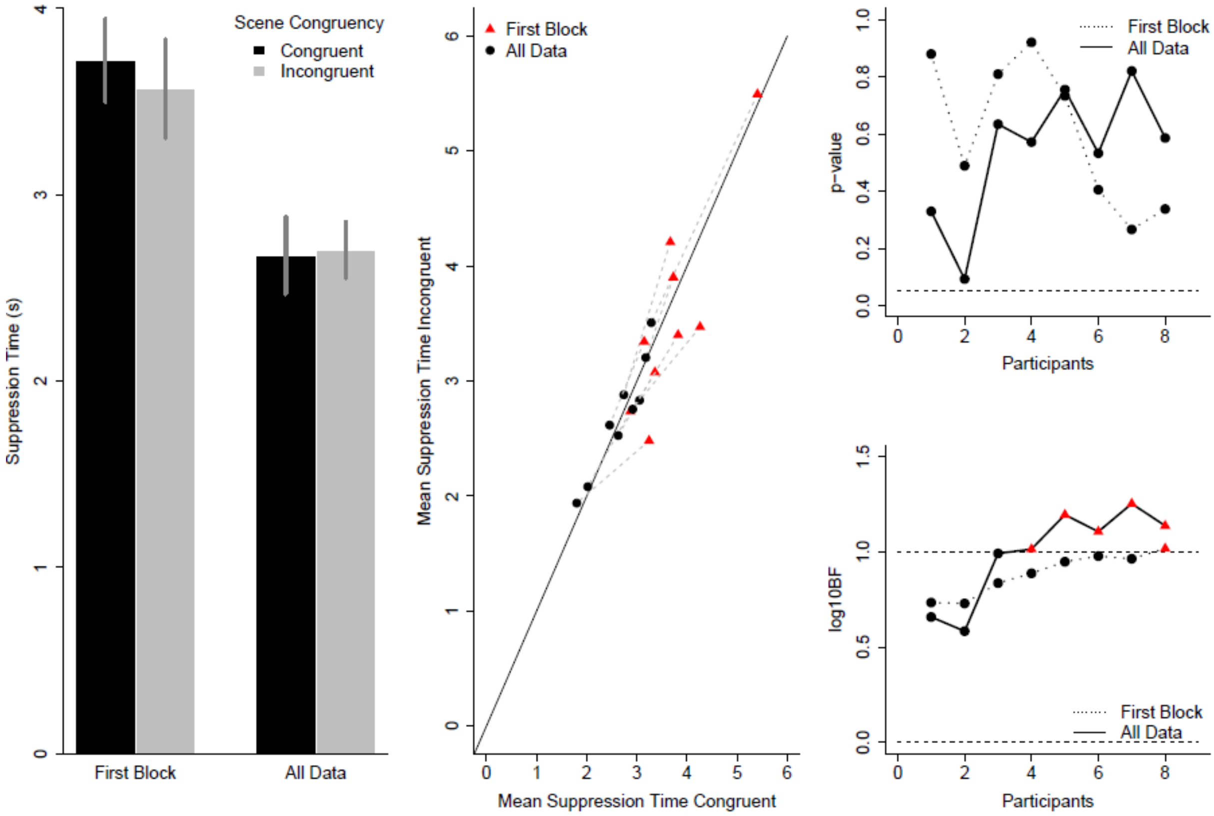

Figure S9. Results of Experiment 2 after removing incorrectly categorized scenes and participants for whom the CFS mask was presented in the nondominant eye. Mean suppression time (a) is graphed for the first block and for all blocks, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown for each participant's performance in the first block and in all blocks. Each dashed line connects the data points for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes. The two other graphs show *p* values (c) and Bayes factor values (d) as a function of the number of participants in the analysis, separately for the first block and for all blocks. Bayes factors (log10) greater than 0 indicate evidence for the absence of an effect, whereas Bayes factors smaller than 0 indicate evidence for the presence of an effect.

Supplementary analysis for Experiment 3

For Experiment 3, we reported an analysis of the mean suppression times after removing incorrect responses during the main experiment as well as incorrectly categorized stimuli during the post-experimental rating session. Here, we report on the same analysis, yet including the stimuli that were incorrectly categorized during the post-experimental rating session. The results of this analysis were very similar to those excluding the incorrectly categorized stimuli (see Figure 10). No effect of scene congruency was observed ($M_{congruent}$ = 2.64, $M_{incongruent}$ = 2.63; $SD_{congruent}$ = 1.12, $SD_{incongruent}$ = 1.06; $t(49)$ = 0.11, p = .91, d = .016). Similarly, the BF analysis indicated convincing evidence for the absence of a congruency effect (BF = 24).

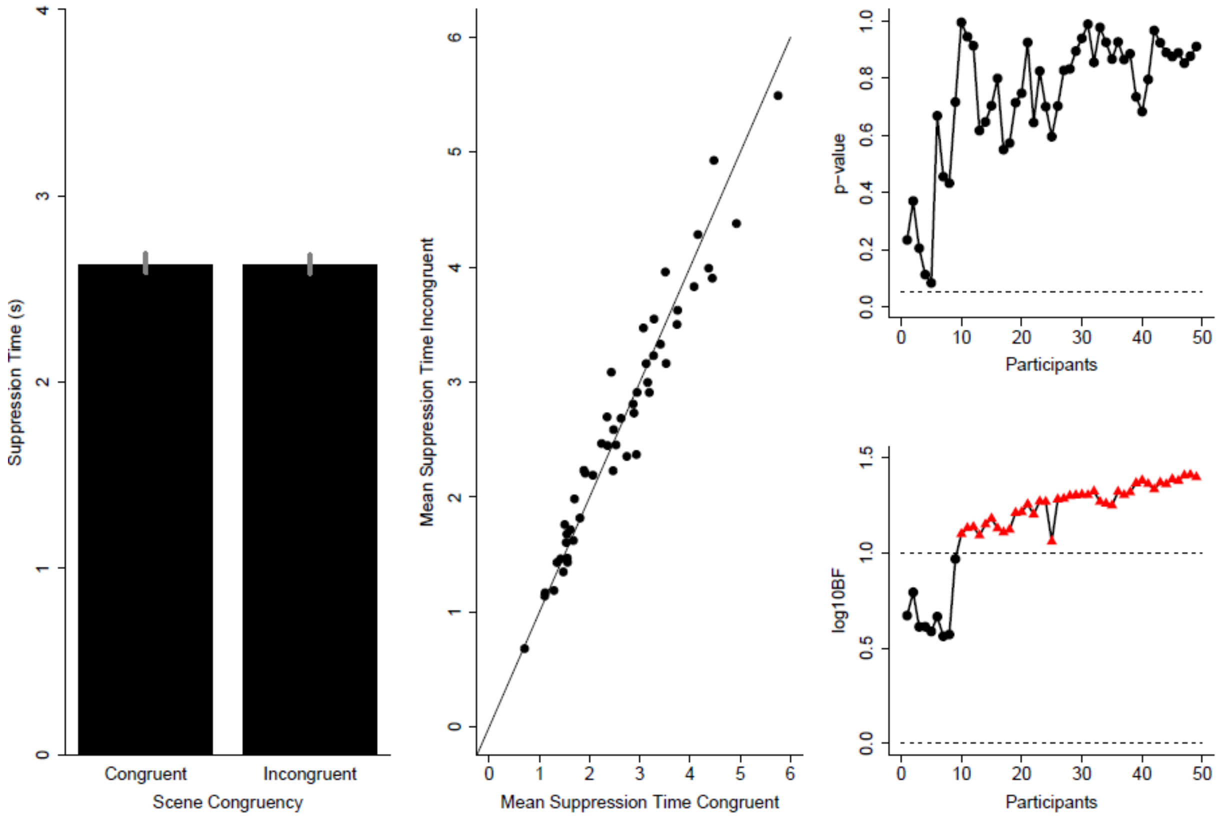

Figure S10. Results of Experiment 3 after including incorrectly categorized scenes. Mean suppression time (a) is graphed separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown. The diagonal line represents equal suppression durations for congruent and incongruent scenes. The two other graphs show p values (c) and Bayes factor values (d) as a function of the number of participants in the analysis. Bayes factors (log10) greater than 0 indicate evidence for the absence of a congruency effect, whereas Bayes factors smaller than 0 indicate evidence for the presence of a congruency effect.

Correlation between suppression times and statistical properties of the images

A final exploratory analysis pertains to the relationship between mean suppression times obtained for each image averaged across observers and four statistical properties of the images. These four statistical measures include the intercept (IC) and slope (SL) derived from a regression line fitted to the Fourier amplitude spectrum of each image in log-log space. Furthermore, for each image, we obtained measures of contrast energy (CE) and spatial coherence (SC) as reported in Groen, Ghebreab, Prins, Lamme, and Scholte (2013). These measures involve approximations of a Weibull fit to the histograms of local contrast filter responses. Note that high values of SC indicate cluttered images, whereas low values indicate images that are spatially coherent.

Figure S11 depicts the correlation matrices for all measures considered and mean suppression time (ST) for each item, averaged across observers and scene congruency. For all experiments, the highest correlation of interest that was observed was the one between suppression time and spatial coherence (Experiment 1: r = 0.45; Experiment 2: r = 0.53; Experiment 3: r = 0.49). This correlation was also the only one (of the ones that included suppression time) that exceeded a Bayes Factor of 3 for all experiments (Wetzels & Wagenmakers, 2012).

This supplementary exploratory analysis indicates that the observed suppression durations correlate with a measure of spatial coherence of the images, indicating that cluttered images on average yield slower suppression times compared to spatially coherent, less fragmented images.

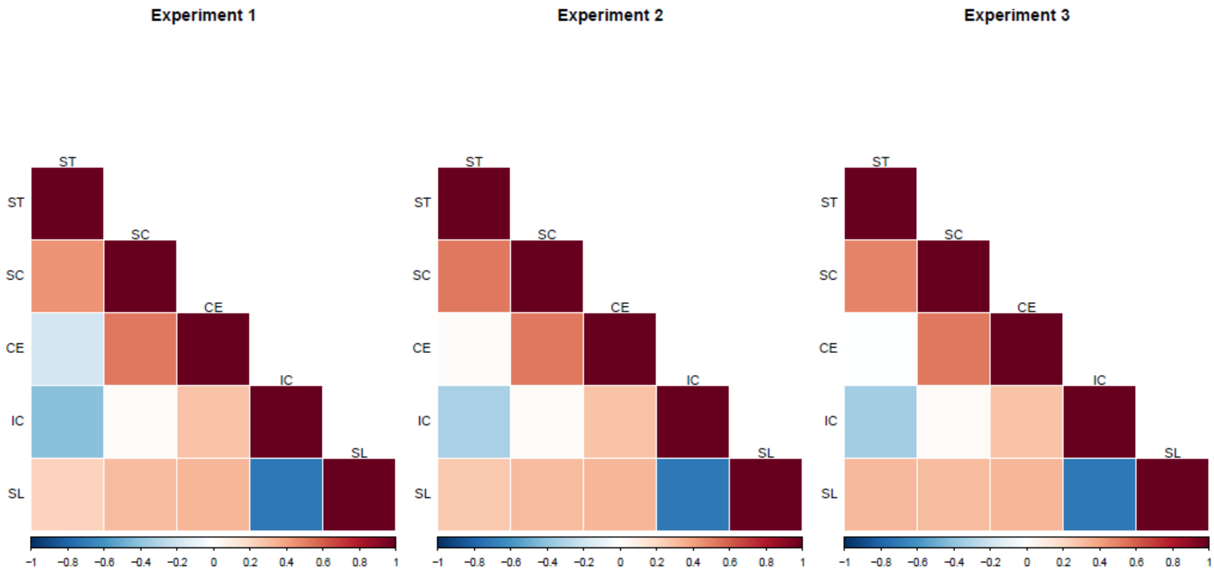

Figure S11. Relationship between suppression times and image characteristics. Correlation matrices for all four measures and suppression time, for both experiments. ST = suppression time, SC = spatial coherence, CE = contrast energy, IC = intercept, SL = slope.

Table S1. Bayes Factors for correlations between suppression time and each image statistic.

Statistic	**Experiment 1**	**Experiment 2**	**Experiment 3**
SC	7.6	53	21
CE	0.23	0.12	0.12
IC	4.2	0.97	1.27
SL	0.32	0.49	0.97

Note. Bayes Factors > 1 indicate evidence for a correlation being different from zero.

REFERENCES

Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V. A. F., & Scholte, H. S. (2013). From Image Statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *The Journal of Neuroscience*, *33*(48), 18814–18824. http://doi.org/10.1523/JNEUROSCI.3128-13.2013

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064. http://doi.org/10.3758/s13423-012-0295-x